1 Redes Convolucionais

Redes convolucionais figuram entre as arquiteturas conhecidas na literatura como pertencentes à uma subárea particular de Redes Neurais , conhecida como *Deep Learning*. *Deep Learning* consiste em um grupo de topologias de redes neurais, o qual tem como grande característica a presença de muitas camadas, apresentando grande aplicação para reconhecimento de padrões. Outra forte característica deste grupo, se encontra no grande espaço amostral exigido para o treinamento das redes, reforçando os obstáculos já conhecidos em outras topologias relacionados à complexidade computacional [1]. Para tornar viável o aprofundamento de camadas, considerando um espaço de treino significativo, fez-se necessário o estudo de implementações alternativas de redes neurais, revolucionando as arquiteturas convencionais. Neste contexto, nascem as redes convolucionais, as quais tem o compromisso de lidar com grande complexidade computacional, não só pelas presença de várias camadas, mas por serem utilizadas majoritariamente para o reconhecimento de imagens [3].

Imagens digitais nada mais são que conjuntos de dados, os quais representam combinações de cores dos *pixels* que as compõem. Uma imagem digital, abstraindo para sua representação numérica discretizada em cores, consiste em uma grande matriz de diversas dimensões, sendo usualmente representada por um volume de dados. Pode-se facilmente compreender a magnitude do esforço computacional exigido para operações de manipulação de imagens, sendo algumas delas: aplicações de filtros; identificação de bordas; transformadas e outras. De nada supreende o fato de que fabricantes de *hardware* vem desenvolvendo alternativas dedicadas ao processamento de imagens, como as GPU *(Graphical Processing Unit)*, as quais possuem uma arquitetura que permite grande poder de paralelização computacional.

Neste cenário, torna-se claro o desafio das redes convolucionais em treinar seus neurônios para reconhecimento de imagens, considerando o volume intenso de dados que elas manipulam. Porém, talvez equiparável ao tamanho do desafio, seja também o investimento para vencê-lo, já que o poder de reconhecimento visual por um sistema inteligente, possui aplicação nas mais diversas indústrias. De automação industrial à segurança, com pouco esforço mental é fácil imaginar uma possível implementação de reconhecimento de imagens interessante para absolutamente qualquer indústria. Não obstante, o pesado investimento [4] em forma de pesquisas e competições acelerou a viabilidade das redes convolucionais quando utilizadas para o propósito em questão. Como grande representativo deste movimento, pode-se citar a competição *ImageNet*, que reúne grandes pólos de tecnologia (iniciativa privada e centros de pesquisa) em prol do desenvolvimento desta área. Entre os notórios participantes desta competição, destacam-se equipes da Microsoft Google, Intel e outras.

Externado o seu posicionamento na sociedade e sua aplicação, faz-se necessário um estudo de sua topologia para entender como redes convolucionais funcionam perante estes desafios.

2 Topologia das Redes Convolucionais

Para melhor entender sua topologia, é de interesse do leitor conhecer de forma macro como uma rede convolucional permite a segmentação de imagens. De forma resumida, a rede convolucional está de forma insistente à procura de representações características de imagens que permitam à ela categorizar a imagem em questão. Esta procura é feita a partir da aplicação de filtros, os quais acusam quando uma representação característica em especial é encontrada na imagem. Como exemplo, pode-se ver na figura abaixo os filtros utilizados para a identificação de uma face e de um carro, aplicados pelas camadas de uma rede convolucional.

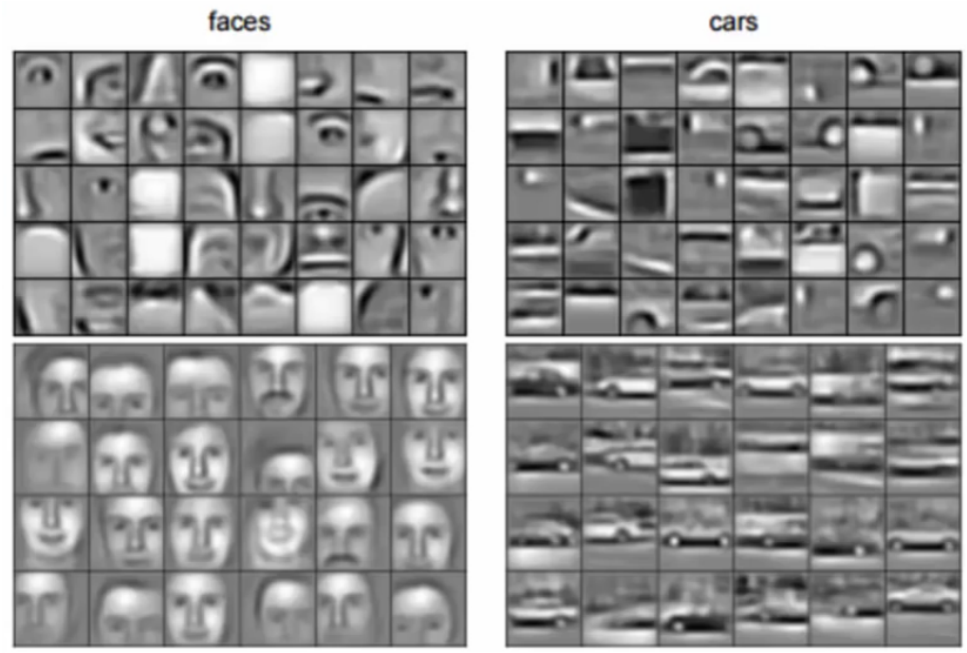

Figura 1: Filtros de uma rede convolucional

A princípio, tendo apenas a figura 1 como referência de filtros utilizados pelas camadas comvolucionais, torna-se difícil entender a praticidade e a real justificativa desta rede, já que olhos, narizes e bocas são obviamente representações características de uma face. Porém, esta primeira impressão é rapidamente questionada ao analisar a figura 2, a qual apresenta filtros que possuem o mesmo propósito.

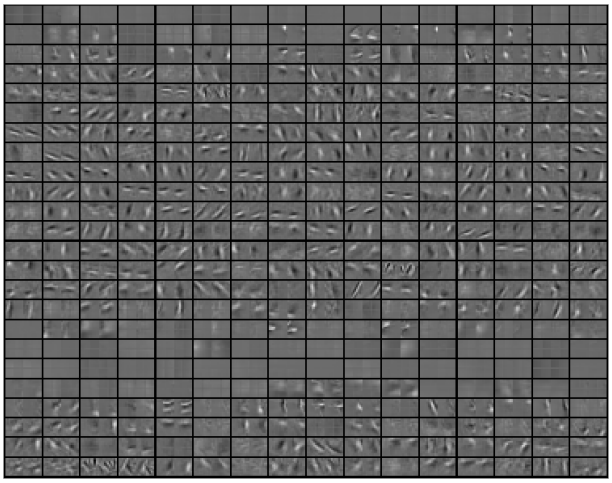

Figura 2: Filtros de uma rede convolucional. Fonte: www.cs.toronto.edu.

Nota-se que estes filtros carregam quase nenhuma semelhança aparente com alguma imagem especial, e visualmente não proporcionam qualquer informação. Este filtros, os quais são frutos do treino da rede, mostram o verdadeiro poder das redes convolucionais, pois um ser humano provavelmente não diria que este filtros tem capacidade alguma de segmentar entre dois ou mais grupos de imagens distintas. O poder de abstração dos filtros em uma rede convolucional vai além da capacidade de segmentação visual do ser humano, sendo então muito mais eficiente que qualquer algoritmo determinístico que possamos desenvolver racionalmente.

A aplicação dos filtros é realizada nas camadas de convolução, sendo esta a principal camada em uma rede convolucional. As demais serão descritas logo a seguir:

2.1 Camada Convolucional

Como explicado anteriormente, a camada convolucional aplica filtros na imagem à procura de sua representação equivalente. Esta aplicação pode ser traduziada matematicamente como a convolução do filtro ao longo da imagem. A figura 3 passa a sensação que o filtro desliza pela imagem [3].

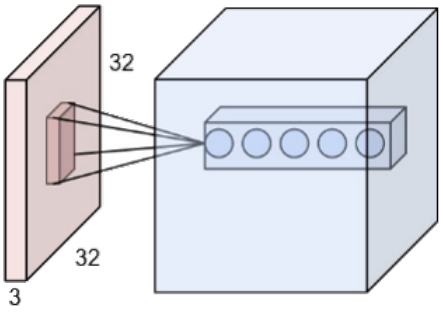

Figura 3: Representação da camada convolucional aplicando filtros em uma imagem. Fonte: [3].

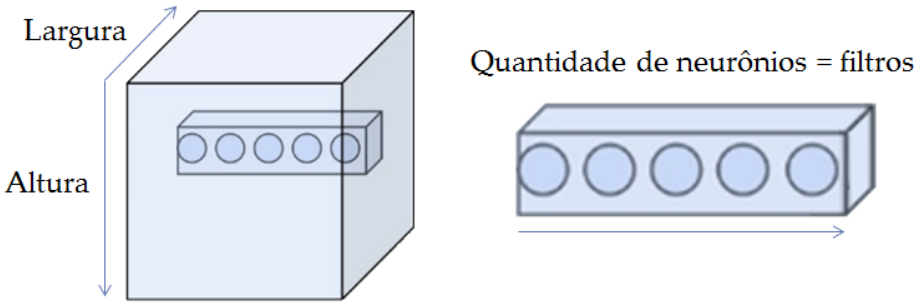

Figura 4: Desenho que retrata a camada convolucional e a disposição de seus neurônios. Fonte: [3].

Na figura 3, o volume rosa representa a imagem alimentada na rede; o volume azul corresponde aos neurônios que estão aplicando o filtro; e a projeção do volume azul corresponde ao filtro sendo aplicado na imagem. Os filtros tem sua representação numérica nos pesos dos neurônios, os quais são calculados durante o treino da rede. Como o mesmo filtro é aplicado ao longo da imagem, e os filtros são os pesos dos neurônios, necessariamente os neurônios que compartilham o mesmo eixo da largura e altura na figura 4, compartilham também os mesmos pesos. Portanto, o número de filtros diferentes aplicados à imagem está relacionado com a quantidade de neurônios no eixo da profundidade (vide 4). A saída dos neurônios quantificam quão próximo uma porção específica da imagem se assemelhou com o fitro aplicado, sendo esta quantificação convencionada como valor de ativação.

O tamanho do filtro, em quantos *pixels* é deslocado o filtro, e outras características importantes da rede convolucional são determinadas pelos seus hiperparâmetros, os quais estão descritos a seguir:

1. F = Tamanho do filtro (FxF).

2. S = *Stride*. Este hiperparâmetro corresponde ao deslocamento do filtro ao longo da imagem, medido em *pixels*.

3. K = Quantidade de filtros. Corresponde à quantidade de neurônios na camada convolucional.

4. W = Resolução da Imagem (WxW).

5. P = *Zero Padding*. Corresponde ao número de zeros adicionados na periferia da imagem. Este hiperparâmetro tem como principal função adequar o tamanho da imagem com o F e S escolhidos.

A partir destes hiperparâmetros, considerando a aplicação do filtro já detalhada anteriormente, os dados de saída possuem resolução (ou área, caso uma notação geométrica esteja em vigor) que pode ser descrita pela fórmula:

$$\text{Resolução} = \frac{W - F + 2P}{S} + 1 \tag{1}$$

Pode-se notar que fixando os hiperparâmetros W, F e S, faz-se necessário uma escolha de P para que a divisão resulte em um número inteiro. A figura 5 mostra um exemplo em que W = 5, F = 3, P = 1 e S = 2:

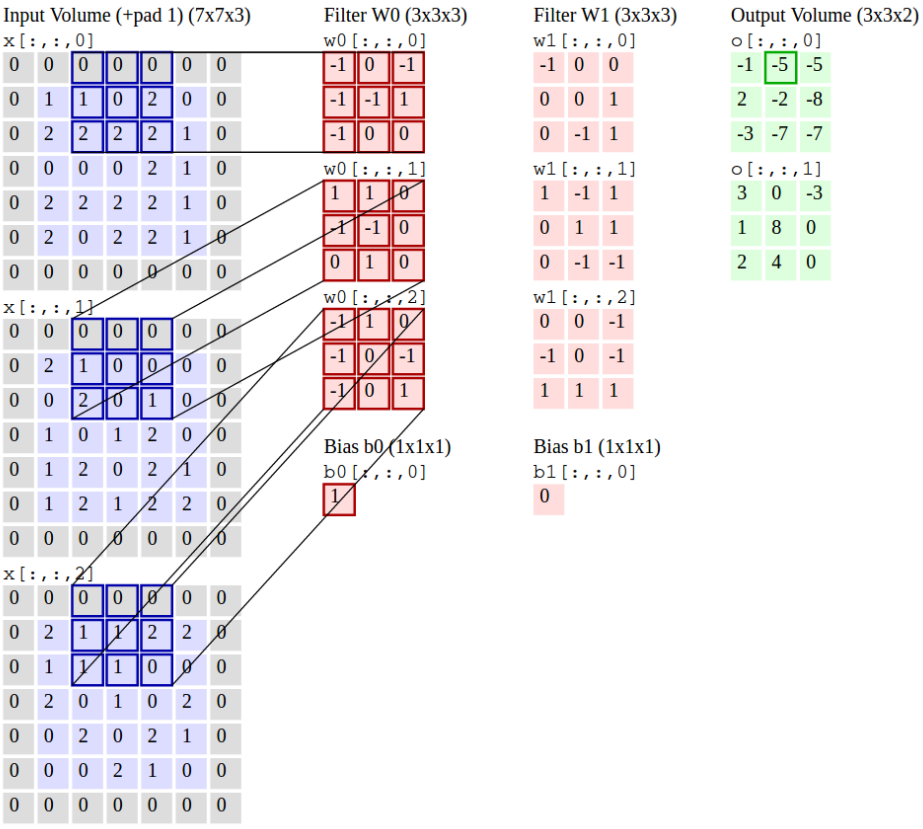

Figura 5: Exemplo de convolução e hiperparâmetros de uma rede convolucional. Fonte: [3].

Não é apenas de camadas convolucionais que uma rede convolucional é feita, outras camadas também realizam operações interessantes que auxiliam na classificação de imagens:

2.2 Camada ReLUs

Esta camada tem uma função muito simples dentro da rede: Rejeitar os baixos valores de ativação, desprezando então as convoluções entre filtro e porção da imagem que não trouxeram informação, e enfatizar os casos contrários. Em muitos casos, as camadas ReLUs aplicam nas saídas do neurônios da camada convolucional, a simples relação matemática:

$$f(x) = max(0, x) \tag{2}$$

Assim, os valores negativos de ativação são desprezados, ou seja, são zerados, passando apenas os que acusaram alguma semelhança entre o filtro e a porção da imagem analisada. Em alguns casos, implementa-se uma função matemática alternativa, conhecidada como

leaky. Ela permite exatamente o que sua tradução em português sugere, um vazamento de valores inferiores à 0. Esta implementação é usada já que a fórmula convencional acaba ocasionalmente neutralizando uma grande sequência de neurônios, já que os valores zerados vão alimentar outras camadas, inativando outros neurônios no caminho. A fórmula alternativa se encontra a seguir:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.01x, & \text{c.c.} \end{cases}$$

Uma camada ReLU pode ser vista em ação na figura 6:

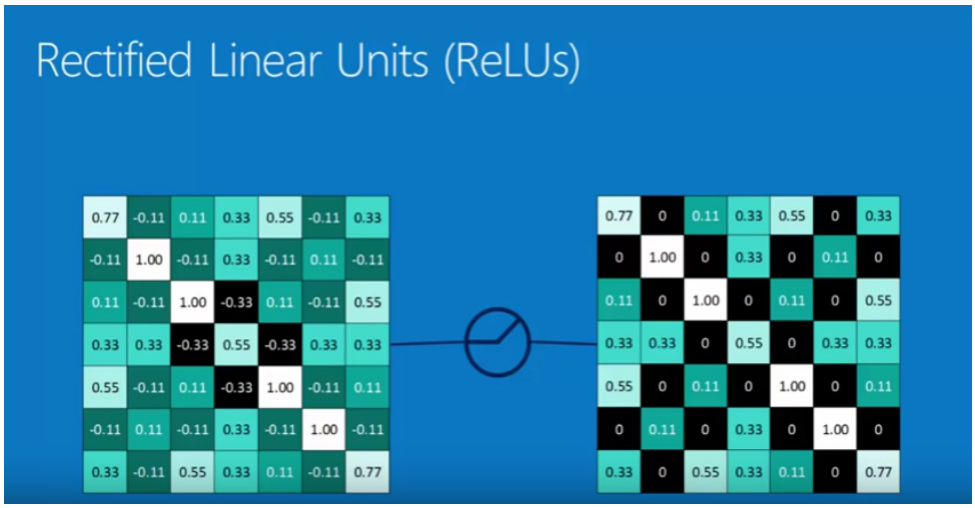

Figura 6: Representação gráfica do funcionamento de uma camada ReLUs em uma rede convolucional. Fonte: https://www.youtube.com/watch?v=FmpDIaiMIeAt=943s.

2.3 Pooling

Esta camada tem como função diminuir o tamanho da imagem ao longo do seu aprofundamento na rede. O seu funcionamento pode ser melhor entendido com a figura 7:

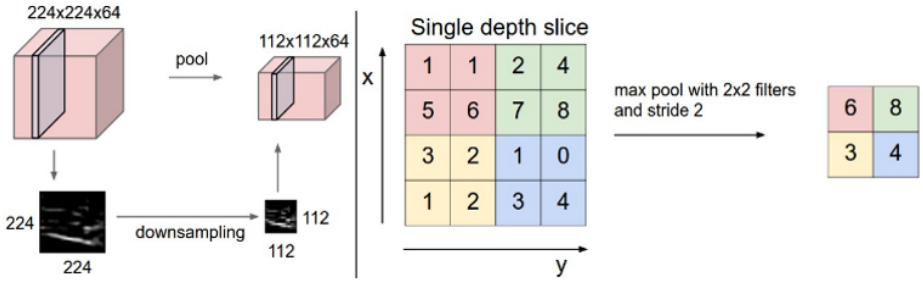

Figura 7: Representação gráfica do funcionamento de uma camada de *Pooling* em uma rede convolucional. Fonte: [3]

Como pode-se ver na figura 7, os valores de ativação menos significativos foram filtrados, e os mais significativos foram rearranjados em uma matriz de resolução inferior. Seu comportamento tem um compromisso parecido com a ReLUs, sendo a associação das duas uma boa estratégia para identificar as ativações que provavelmente carregam informações valiosas na categorização da imagem.

2.4 MLP - *Multilayer Perceptron*

Com a alternância sistemática das camadas anteriores, ou seja, gradualmente diminuindo o volume dos dados de entrada, aplicando filtros e coletando seus valores de ativação significativos, ao final das diversas camadas, resta-se apenas um vetor. Este vetor corresponde à uma votação ponderada dos grupos os quais se deseja classificar as imagens. Portanto, se uma classificação está sendo realizada para separar veículos de trasnporte, o vetor de votação pode estar sugerindo que a imagem alimentada na rede é provavelmente um carro. Para acontecer esta decisão, a camada final de uma rede convolucional consiste geralmente em uma rede MLP (*MultiLayer Perceptron*) convencional, em que cada neurônio se conecta com todos os demais na camada seguinte. A figura 8 relembra o que seria uma topologia MLP:

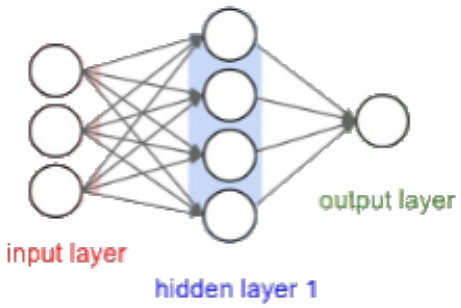

Figura 8: Topologia de uma rede MLP. Fonte: [3].

Com todas as camadas descritas, pode-se entender finalmente como funcionaria uma implementação típica de uma rede convolucional para a classificação de imagens. A figura 9 mostra uma rede convolucional típica em ação:

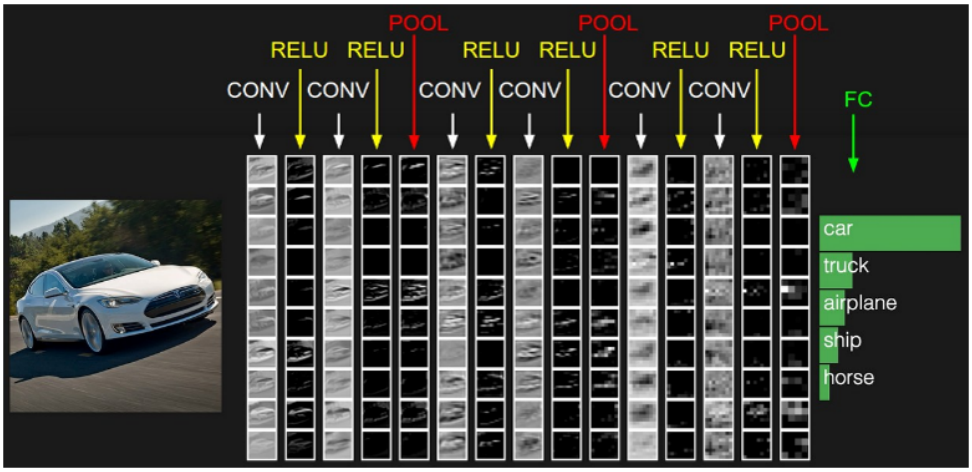

Figura 9: Aplicação típica de uma rede convolucional para a classificação de imagens. Fonte: [3].

Referências

[1] H. Simon, *Redes Neurais: princípios e práticas*, 2.ed, Porto Alegre: Bookman 2001.

[2] J. Hertz et al, *Introduction to the Theory of Neural Computation*, A Lecture Notes Volume in the Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company, California, 1993.

[3] J. Johnson et al *CS231n: Convolutional Neural Networks for Visual Recognition*, Lecture Notes http://cs231n.stanford.edu/.

[4] Investments in Image Recognition *Index*, https://index.co/market/image-recognition/investments.